Fiche TD2 : Langages

Exercice 1 (*) *Parties de mots*

1. Enumérer les suffixes communs aux mots suivants : nation, donation, obstination. Proposer d'autres mots de la langue française ayant parmi leurs suffixes tous ceux précédemment cités.

2. On considère le mot m suivant : embarrassante.

   a) Donner un préfixe propre de m qui soit un mot de la langue française.

   b) Donner un facteur de m (ni préfixe, ni suffixe) qui soit un mot de la langue française.

   c) Donner trois mots de la langue française qui sont des sous-mots de m sans en être facteur.

Exercice 2 (**) *Commutativité de mots*

L'objectif de cet exercice est de donner un critère nécessaire et suffisant pour que deux mots commutent. On se donne un alphabet fini Σ ; dans la suite, "mot" signifie par défaut "mot de $\Sigma^\star$".

1. Soit x et y deux préfixes d'un même mot m. Montrer que x est préfixe de y ou y est préfixe de x.

2. Soient u, v, w, z quatre mots de tels que $uv = wz$. Montrer qu'il existe un unique mot t tel que l'une des conditions suivantes est vérifiée :

   - $u = wt$ et $z = tv$
   - $w = ut$ et $v = tz$

   Remarque : Ce résultat porte le nom de lemme de Lévi et se généralise dans certaines familles de monoïdes.

3. En déduire que deux mots commutent pour la concaténation si et seulement si ils sont puissances d'un même mot. Reformulé, il s'agit de montrer que

$$\forall\, u, v \in \Sigma^\star,\ uv = vu \text{ si et seulement si } \exists\, w \in \Sigma^\star, \exists\, (\alpha, \beta) \in \mathbb{N}^2, u = w^\alpha \text{ et } v = w^\beta$$

   Indication : Pour le sens le plus difficile, on pourra procéder par récurrence sur $|uv|$.

Pour tout mot $m = m_0 ... m_n \in \Sigma^\star$, on note ${}^t m$ le mot $m_n ... m_0$. Le langage $\mathrm{Pal}(\Sigma^\star)$ des palindromes sur $\Sigma^\star$ est défini comme suit : $\mathrm{Pal}(\Sigma^\star) = \{m \in \Sigma^\star \mid m = {}^t m\}$.

4. (***) Donner une condition nécessaire et suffisante (et prouver qu'elle l'est) simple pour que uv soit un palindrome lorsque les mots u et v en sont.

Exercice 3 (**) *Mots et langage de Dyck*

Dans cet exercice, on travaille sur l'alphabet $\Sigma = \{a, b\}$. On introduit une valuation ν sur Σ telle que $\nu(a) = 1$ et $\nu(b) = -1$ qu'on étend aux mots de $\Sigma^\star$ de la façon suivante : pour tout mot $m = m_0 ... m_n \in \Sigma^\star$, $\nu(m) = \sum_{i=0}^{n} \nu(m_i)$. Un mot $m \in \Sigma^\star$ est un mot de Dyck (on dit aussi, "mot bien parenthésé" pour des raisons qui devraient être évidentes vu leur définition) si $\nu(m) = 0$ et tout préfixe p de m vérifie $\nu(p) \geq 0$. On note $\mathcal{D}$ l'ensemble des mots de Dyck et on l'appelle le langage de Dyck.

1. Donner quelques exemples de mots de Dyck et de mots de $\Sigma^\star$ qui ne sont pas des mots de Dyck.

2. Montrer que $\varepsilon \in \mathcal{D}$ et que pour tous mots $u, v \in \mathcal{D}$, le mot $aubv$ est aussi dans $\mathcal{D}$.

3. Montrer que tout mot de Dyck $m \neq \varepsilon$ se décompose sous la forme $aubv$ avec $u, v \in \mathcal{D}$.

   Indication : Considérer le plus petit préfixe non vide de m qui a une valuation nulle.

4. Montrer que la décomposition précédente est unique.

Les deux questions précédentes donnent une caractérisation des mots de Dyck qui permet de les dénombrer :

5. Pour tout $n \geq 0$, on note C_n le nombre de mots de Dyck de taille $2n$. Montrer que la suite $(C_n)_{n\in\mathbb{N}}$ vérifie :
$$C_0 = 1 \text{ et pour tout } n \geq 1,\ C_n = \sum_{k=0}^{n-1} C_k C_{n-1-k}$$

6. En déduire que pour tout $n \geq 0$ le nombre de mots de Dyck de taille $2n$ est $\frac{1}{n+1}\binom{2n}{n}$.

Remarque : Nous avons déjà rencontré la suite $(C_n)_{n\in\mathbb{N}}$ l'année dernière : il s'agit de la suite de Catalan. En effet, C_n est non seulement le nombre de mots bien parenthésés à $2n$ parenthèses mais aussi le nombre de squelettes d'arbres binaires à n noeuds. De manière générale, cette suite intervient fréquemment en combinatoire.

Exercice 4 (***) *Résiduels d'un langage*

Soit Σ un alphabet fini. Si m est un mot de $\Sigma^\star$ et L est un langage sur $\Sigma^\star$, on appelle *résiduel de L par rapport à m* le langage $m^{-1}L = \{u \in \Sigma^\star \mid mu \in L\}$. L'ensemble $\{m^{-1}L \mid m \in \Sigma^\star\}$ s'appelle *l'ensemble des résiduels de L*.

1. Dans cette question, on note L le langage dénoté par l'expression rationnelle $b^\star ab^\star$ sur l'alphabet $\{a, b\}$.
   a) Calculer le résiduel de L par rapport à a (c'est-à-dire, donner une caractérisation du langage $a^{-1}L$).
   b) Reprendre la question précédente en changeant a par ab. Que constate-t-on ?
   c) Montrer que l'ensemble des résiduels de L est fini et le calculer.
2. Dans cette question, on considère le langage $L = \{u^2 \mid u \in \{a, b\}^\star\}$. Montrer que L admet une infinité de résiduels. *Indication : Considérer les résiduels de L par rapport à chacun des mots ba^n où $n \geq 0$.*

Remarque (qui s'éclairera après le chapitre sur les automates) : On peut montrer qu'un langage est rationnel si et seulement si son nombre de résiduels est fini - ce qui donne une façon de prouver le caractère rationnel d'un langage. Il existe d'autre part un lien fort entre le nombre de résiduels d'un langage rationnel L et le nombre minimal d'états d'un automate permettant de reconnaître L.

Exercice 5(*) *Mots dans un langage rationnel*

On se place sur l'alphabet $\Sigma = \{A, T, C, G\}$. Dans chacun des cas suivants, déterminer tous les mots de longueur au plus 3 qui appartiennent au langage dénoté l'expression rationnelle considérée :

1. $(T + TA)^\star$
2. $AC^\star + C$
3. $(G + \varepsilon)TT^\star$
4. $(A + C)^\star ACC$
5. $A^\star(C + G)T^\star$
6. $(CG + \varepsilon)G$

Exercice 6 (*) *Du langage rationnel à sa description*

On se place sur l'alphabet à deux lettre $\Sigma = \{a, b\}$. Donner une description simple en français des langages dénotés par les expressions régulières suivantes :

1. $(\varepsilon + \Sigma)(\varepsilon + \Sigma)$
2. $(\Sigma^2)^\star$
3. $(b + ab)^\star(a + \varepsilon)$
4. $(ab^\star a + b)^\star$
5. $a^\star(a + b)^\star$
6. $(a^\star b^\star)^\star$

Que peut-on dire des expressions rationnelles 5 et 6 ? Lesquelles parmi ces expressions rationnelles sont linéaires ?

Exercice 7 (*) *De sa description au langage rationnel*

Dans chacun des cas suivants, déterminer une expression rationnelle qui dénote le langage décrit :

1. Les mots sur l'alphabet $\{0, 1\}$ qui contiennent au moins un 0.
2. Les mots sur l'alphabet $\{0, 1\}$ qui contiennent au plus un 0.

3. Les mots sur l'alphabet $\{0, 1, 2\}$ qui commencent par 0, finissent par 1 et contiennent au moins un 2.
4. Les mots sur l'alphabet $\{0, 1\}$ tels que toute série de 0 consécutifs est de longueur paire.
5. Les mots sur l'alphabet $\{0, 1\}$ tels que deux lettres consécutives sont toujours distinctes.
6. Les mots sur un alphabet quelconque Σ dont la longueur n'est pas divisible par 3.

Exercice 8 (**) *Vrai ou faux ?*

Dans la suite, a et b sont des lettres d'un alphabet. Pour chacune des affirmations suivantes, déterminer si elle est vraie ou fausse en justifiant :

1. Les langages $a^\star b^\star$ et $\{a, b\}^\star$ sont égaux.
2. Les langages $(a^\star + b^\star)^\star$ et $\{a, b\}^\star$ sont égaux.
3. L'assertion précédente reste vraie si on remplace a et b par n'importe quelles expressions rationnelles.
4. Le langage $a^\star bc^\star$ est local.
5. Le langage $a^\star ba^\star$ est local.
6. Un langage rationnel est local.
7. Tout langage rationnel est infini.
8. Il y a une infinité de langages non rationnels.
9. Tout langage inclus dans un langage rationnel est rationnel.
10. Tout langage rationnel est dénoté par une infinité d'expressions rationnelles différentes.

Exercice 9 (**) *Stabilités de langages locaux*

1. L'intersection de deux langages locaux est-il un langage local ? Justifier.
2. Le complémentaire d'un langage local est-il un langage local ? Justifier.
3. Si L est un langage local, montrer que le langage L' de ses facteurs est aussi un langage local.